

09-07-00

A

Customer No. 20350
TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200

09/06/00

Attorney Docket No. 023070-079820US
"Express Mail" Label No. EL624021678US
Date of Deposit: September 6, 2000

ASSISTANT COMMISSIONER FOR PATENTS
BOX PATENT APPLICATION
Washington, D.C. 20231

Sir:

Transmitted herewith for filing under 37 CFR 1.53(b) is the patent
application of inventors:

ETZLER, Marilynn E. and ROBERTS, Nicholas J.

jc939 U.S. PTO
09/06/00
09/06/00

For: LNP, A PROTEIN INVOLVED IN THE INITIATION OF MYCORRHIZAL INFECTION IN PLANTS

Enclosed are:

[X] 1 Return Mailroom Postcard Receipt
[X] 1 Transmittal Letter (in triplicate)
[X] 1 Application Cover Sheet
[X] 26 page(s) of specification
[X] 2 page(s) of claims
[X] 1 page of Abstract
[X] 8 sheet(s) of Sequence Listing.

	(Col. 1)	(Col. 2)
FOR:	NO. FILED	NO. EXTRA
BASIC FEE		
TOTAL CLAIMS	14- 20	= *0
INDEP. CLAIMS	1- 3	= *0
[] MULTIPLE DEPENDENT CLAIM PRESENTED		

* If the difference in Col. 1 is less than 0, enter "0" in Col. 2.

SMALL ENTITY

RATE	FEE
	\$345.00
x \$9.00 =	\$0.00
x \$39.00 =	\$0.00
+ \$130.00 =	
TOTAL	\$345.00

OTHER THAN SMALL ENTITY

RATE	FEE
	\$690.00
x \$18.00 =	
x \$78.00 =	
+ \$260.00 =	
TOTAL	

Please charge Deposit Account No. 20-1430 as follows:

[X] Filing fee

\$ 345.00

[X] Any additional fees associated with this paper or during the pendency of this application.

[] A check for \$ _____ is enclosed.
2 extra copies of this sheet are enclosed.

Respectfully submitted,
TOWNSEND and TOWNSEND and CREW LLP

Kevin Bastian

Kevin Bastian
Reg No.: 34,774
Attorneys for Applicant

Telephone:
(415) 576-0200

Facsimile:
(415) 576-0300

Express Mail No.: EL624021678US
Attorney Docket No.: 23070-079820US

PATENT APPLICATION
LNP, A PROTEIN INVOLVED IN THE INITIATION OF
MYCORRHIZAL INFECTION IN PLANTS

Inventor(s):

Marilynn E. Etzler

Nicholas J. Roberts

Assignee:

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
1111 Franklin Street, 5th Floor
Oakland, CA 94607-5200

Entity: Small

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

009050 "TE926960

LNP, A PROTEIN INVOLVED IN THE INITIATION OF MYCORRHIZAL INFECTION IN PLANTS

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application is related to USSN 08/907,226, filed August 6, 1997, and USSN 08/129,112, August 4, 1998, both of which are incorporated herein by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with Government support under Grant No. GM21882, awarded by the National Institutes of Health, under Grant No. DCB 9004967, awarded by the National Science Foundation and under NRICGP Grant number 97-35305-4630 from the United States Department of Agriculture. The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Mycorrhizal associations are structures formed between symbiotic soil fungi and plant roots. Mycorrhizal fungi infect plant roots and enhance the plant's ability to take up water and nutrients, particularly phosphorus, from the soil. As obligate symbionts, mycorrhizal fungi are unable to live outside of a plant host.

It is estimated that more than 80% of flowering plant species on land are able to form mycorrhizal associations. Mycorrhizae fall into two groups based on the interacting plant hosts and fungal species. Woody Angiosperms and Gymnosperms interact with the fungi Basidiomycetes, Ascomycetes or Zygomycetes to form ectomycorrhizae. The Zygomycetes form endomycorrhizae with most other terrestrial plant species. Arbuscular mycorrhizae, an endomycorriza, are the most common form of mycorrhizae. Arbuscular mycorrhizal fungi interact nonspecifically with plants, and a single fungal species can symbiose with many plant species. (Gianinazzi-Pearson, V. The Plant Cell 8:1871-83 (1996); Harrison, M.J. Trends in Plant Science 2:54-60 (1997)). However, twenty percent of plants on land are unable to form or rarely form mycorrhizal associations, including the families Brassicaceae, Cyperaceae, Cruciferae, Chenopodiaceae, and Caryophyllaceae. (Raven, P. H., et al., BIOLOGY OF PLANTS. p.224, Worth Publishers, New York (1989); Sharma, S. et al.. Microbiologia Sem. 13:427-436 (1997)).

009060" TEL 2960

Mycorrhizal associations provide a number of benefits to the host plant in addition to the well-documented enhancement of phosphate uptake. Mycorrhizal associations have been shown to stimulate uptake of nitrogen, zinc, copper, sulfur, potassium, and calcium and to enhance the uptake of water. Mycorrhizal associations also protect the plant host from infection by pathogens. (Sharma, S. et al., Microbiologia Sem. 13:427-436 (1997)).

After infection with mycorrhizal fungi, crop plants exhibit improved growth. Economically important plants tested include vegetables, and field crops. See references in U.S. Patent 5,096,481. Mycorrhizal fungal infection also enhances plant growth under stress conditions, including growth on reclaimed soils.

Previously, a root lectin, LNP (formerly called NBP46 or DB46) was isolated from young *Dolichos biflorus* root extracts. LNP is a 46 kDa protein that was isolated by affinity chromatography on hog gastric mucin blood group A + H substance conjugated to Sepharose (Quinn, J.M. and Etzler, M.E. Arch. Biochem. Biophys. 258:535-544 (1987)). The protein also has apyrase activity and appears to play an early role in rhizobium-legume biosynthesis.(Etzler, M.E. et al., Proc. Natl. Acad. Sci. USA 96:5856-5861 (1999)). Genetic experiments indicate that the establishment of rhizobial symbiosis and mycorrhizal symbiosis share common steps (Albrecht, C. et al., EMBO J. 18:281-288 (1999)).

Identification of genes and proteins that modulate mycorrhizal fungal association with plants will further the beneficial use of mycorrhizae in agriculture. For example, plants that have enhanced mycorrhizal association will be able to use nutrients more efficiently and potentially require less fertilizer or be able to grow on less fertile soil. Plants that currently do not associate with mycorrhizal fungi could be transformed with genes to allow the association to take place, thereby increasing the range of environments for growth of these plants.

One of the obstacles to greater use of mycorrhizal fungi in agriculture is the difficulty in growing large quantities of the fungus. All mycorrhizal fungi are obligate symbionts and cannot be grown outside of the plant. Innoculants are most commonly made from the roots of infected plants. Enhancement of mycorrhizal fungal infection will improve the yield and efficacy of inoculant stocks. The present invention addresses these and other needs.

SUMMARY OF THE INVENTION

The present invention provides methods of modulating mycorrhizal infection in a plants. The method comprise introducing into the plant an expression cassette containing a plant promoter operably linked to a heterologous LNP polynucleotide or complement thereof,

wherein the LNP polynucleotide encodes an LNP polypeptide at least about 70% identical to SEQ ID NO:2, SEQ ID NO: 4, or SEQ ID NO: 6. The heterologous LNP polynucleotide can be SEQ ID NO:1, SEQ ID NO: 3, or SEQ ID NO: 5.

The expression cassette can be introduced into the plant using any standard technique, including *Agrobacterium*-mediated transformation or through a sexual cross. In the expression cassette, the promoter can be linked to the *LNP* polynucleotide in an antisense or sense orientation.

Typically the methods are used to enhance expression of the *LNP* polynucleotide, thereby increasing infection of the plant by a mycorrhizal fungus. The method may further comprise infecting the plant with a mycorrhizal fungus, such as *Glomus intraradices*.

DETAILED DESCRIPTION OF THE INVENTION

1. Definitions

The phrase “mycorrhizal infection” refers to all interactions between mycorrhizal fungus and a transgenic plant. The term includes, but is not limited to, recognition between a plant polypeptide and structures on the fungus, binding of mycorrhizal fungi to the plant root, the symbiotic relationship between a mycorrhizal fungus and a plant, and mycorrhizal fungus-plant interactions, including any morphological or molecular changes initiated on interaction between the plant and the mycorrhizal fungus. The morphological changes can occur in either the plant or the mycorrhizal fungi and include specialized structures to facilitate exchange of nutrients between the plant and the fungi. For example, some species of mycorrhizal fungi form arbuscules and vesicles within plant cells. The molecular changes include, but are not limited to, signal transduction cascades that result in changes in expression of gene products. The changes of gene product expression may occur either transcriptionally or post-transcriptionally.

The phrase “isolated nucleic acid molecule” or “isolated protein” refers to a nucleic acid or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *LNP* gene is separated from open reading frames which flank the gene and encode a protein other than LNP. The term

"purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an R1 generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

The phrase "operably linked" refers to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "polynucleotide," "polynucleotide sequence" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular LNP nucleic acid sequence of this invention also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol et al., 1992; Rossolini et al., *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

An "LNP polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of about 100 to about 2000 nucleotides, sometimes from about 1400 to about 1500 nucleotides, which specifically hybridizes, to the *Dolichos biflorus* polynucleotide (SEQ ID NO:1), or to the *Lotus japonicus* polynucleotide (SEQ ID NO:3), or to the *Medicago sativa* polynucleotide (SEQ ID NO:5), or which encodes an LNP polypeptide. The isolation and characterization of the *Lotus* and *Medicago* genes are described in the PCT application WO 98/16261.

An LNP polypeptide of the present invention comprises at least 50 amino acids, more preferably at least 100 amino acids, still more preferably at least 200 amino acids and most preferably up to about 500 amino acids from SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6, and conservatively modified variants thereof. The LNP polypeptides of the present invention also include proteins which have substantial identity to an LNP protein of at least 10 to 500 amino acids selected from SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6 and conservatively modified variants thereof.

The term "sexual reproduction" refers to the fusion of gametes to produce seed by pollination. A "sexual cross" is pollination of one plant by another. "Selfing" is the production of seed by self-pollonization, i.e., pollen and ovule are from the same plant.

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by the term *LNP* nucleic acid.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*LNP* nucleic acid". In addition, the term specifically includes those sequences substantially identical (determined as described below) with an *LNP* polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type *LNP* polypeptides or retain the function of the *LNP* polypeptide (e.g., resulting from conservative substitutions of amino acids in the *LNP* polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller,

Computer Applic. Biol. Sci. 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to sequences or subsequences that have at least 60% identity. Alternatively, percent identity can be any integer from 60% to 100%, e.g. 60%, 61%, 62%, 63%, ect. More preferred embodiments include at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% nucleotide or amino acid residue identity when aligned for maximum correspondence over a comparison window as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This definition also refers to the complement of a test sequence, which has substantial sequence or subsequence complementarity when the test sequence has substantial identity to a reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing

the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al, supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);

- 3) Asparagine (N), Glutamine (Q);
 - 4) Arginine (R), Lysine (K);
 - 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
 - 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (see, e.g., Creighton, *Proteins* (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30 °C below the T_m . The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent

conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising LNP nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

The phrase "transgenic plant" refers to a plant into which heterologous polynucleotides have been introduced by any means other than sexual cross or selfing. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, in planta techniques, and the like. Such a plant containing the heterologous polynucleotides is referred to here as an R1 generation transgenic plant. Transgenic plants may also arise from sexual cross or by selfing of transgenic plants into which heterologous polynucleotides have been introduced.

2. Introduction

The present invention provides polynucleotides referred to here as *LNP* polynucleotides, as exemplified by SEQ ID NO:1. Polypeptides encoded by the genes of the

invention are lectins involved in binding a variety of carbohydrates. In addition, polypeptides function as an enzyme, catalyzing the dephosphorylation of nucleotide di- and triphosphates.

The polypeptides of the invention are also involved oligosaccharide signaling events that play important roles in the regulation of plant development, defense, and other interactions of plants with the environment. Although the structures of some of these oligosaccharides have been characterized in the prior art, little is known about the plant receptors for these signals, nor the mechanism(s) by which these signals are transduced.

Without wishing to be bound by theory, it is believed that polypeptides of the LNP protein modulate oligosaccharide signaling events that are important for the interaction between mycorrhizal fungi and plants. Mycorrhizal associations are known to enhance phosphate uptake by plants. The LNP protein also has apyrase activity. Thus, LNP may also be involved in the enhancement of phosphate uptake after infection by mycorrhizal fungi.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook, et al.

3. Isolation Of Nucleic Acid Sequences From Plants

The isolation of sequences from the genes of the invention may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the nucleic acid and peptide sequences disclosed herein can be used to identify the desired gene in a cDNA or genomic DNA library from a desired plant species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g., using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a library of tissue-specific cDNAs, mRNA is isolated from tissues and a cDNA library which contains the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned gene such as the polynucleotides disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other in vitro amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes known to those of skill.

Appropriate primers and probes for identifying LNP genes from *Dolichos biflorus* or transgenic plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990). Appropriate degenerate primers for this invention include, for instance: a 5' PCR primer [5'-TA(T/C)GCNGTNAT(T/C)TT(T/C)GATGC-3'] (SEQ ID NO:4) and a 3' PCR primer [5'-AT(A/G)TT(A/G)TA(T/A/G)AT(G/A)CCNGG-3'] (SEQ ID NO:5) where N denotes all nucleotides. The amplification conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μ M dATP, 200 μ M dCTP, 200 μ M dGTP, 200 μ M dTTP, 0.4 μ M primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec, followed by 72°C for 5 min.

Using the above primers, a partial coding sequence will be obtained. There are many techniques known to those of skill to determine and isolate the complete coding sequence. These methods include using the PCR amplified subsequence to probe a cDNA library for longer sequences.

A preferred method is RACE (Frohman, et. al., *Proc. Nat'l. Acad. Sci. USA* 85:8998 (1988)). Briefly, this technique involves using PCR to amplify a DNA sequence using a random 5' primer and a defined 3' primer, e.g., (SEQ ID NO:6) (5' RACE) or a random 3' primer and a defined 5' primer, e.g., (SEQ ID NO:7) (3' RACE). The amplified sequence is then subcloned into a vector where it is then sequenced using standard techniques. Kits to perform RACE are commercially available (e.g. 5' RACE System, GIBCO BRL, Grand Island, New York, USA). In this manner, the entire LNP coding sequence of about 1600 bp can be obtained (SEQ ID NO:1). The invention also provides genomic sequence of the LNP (SEQ ID NO:3).

Alternatively, primers can be selected and synthesized by those of skill from the cDNA sequence disclosed in SEQ ID NOs:1 and 3.

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers, et al., *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams, et al., *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

4. Use Of Nucleic Acids Of The Invention To Modulate Gene Expression

The polynucleotides of the invention can be used to enhance expression (i.e., increase expression of an endogenous gene or provide *LNP* expression in a plant that does not normally express *LNP*) of genes of the invention and thereby enhance infection of transgenic plants by mycorrhizal fungi, increase the level of nutrients taken up by the plants, and affect the growth and development of transgenic plants. Alternatively, enhanced expression can be used to modulate oligosaccharide signaling in the plant. This can be accomplished by the overexpression of LNP polypeptides in the tissues of transgenic plants.

The heterologous *LNP* polynucleotides do not have to code for exact copies of the LNP proteins exemplified herein. Modified LNP polypeptide chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described for instance, in Sambrook et al., *supra*. Hydroxylamine can also be used to introduce single base mutations into the coding region of the gene (Sikorski, et al., *Meth. Enzymol.* 194: 302-318 (1991)). For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Alternatively, the nucleic acid sequences of the invention can be used to inhibit expression of an endogenous gene. One of skill will recognize that a number of methods can be used to inactivate or suppress LNP activity or gene expression. The control of the expression can be achieved by introducing mutations into the gene or using recombinant DNA techniques. These techniques are generally well known to one of skill and are discussed briefly below.

Methods for introducing a genetic mutations into a plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used. Desired mutants are selected by assaying for increased seed mass, oil content and other properties.

Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences. LNP mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of LNP mRNA, e.g., by Northern blots. Mutants can also be selected by assaying for increased seed mass, oil content and other properties.

The isolated sequences prepared as described herein, can also be used in a number of techniques to suppress endogenous LNP gene expression. A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt et al., U.S. Patent No. 4,801,340.

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous *LNP* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least

about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 1700 nucleotides is especially preferred.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *LNP* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli et al., *The Plant Cell* 2:279-289 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184. In addition, a combination of simultaneous sense and antisense expression can also be used for gene silencing in plants (Waterhouse *et al. Proc. Natl. Acad. Sci. U.S.A.* 95: 13959-13964 (1998).

The suppressive effect may occur where the introduced sequence contains no coding sequence per se, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used.

A. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, for example, Weising, et al., *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length LNP protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transgenic plant, i.e., a root-specific promoter.

Promoters can be identified by analyzing the 5' sequences of a genomic clone in which naturally occurring lectin nucleotide phosphohydrolase genes, i.e., *LNP*, can be found. At the 5' end of the coding sequence, nucleotide sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing, et al., in *GENETIC ENGINEERING IN PLANTS*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, et al., *Plant Cell* 1:855-866 (1989); Bustos, et al., *Plant Cell* 1:839-854 (1989); Green, et al., *EMBO J.* 7:4035-4044 (1988); Meier, et al., *Plant Cell* 3:309-316 (1991); and Zhang, et al., *Plant Physiology* 110:1069-1079 (1996)).

009060" 7292560

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter may direct expression of the polynucleotide of the instant invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues, such as roots, fruit, seeds, or flowers. Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (e.g., promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta. Other markers such as green fluorescence protein (GFP), GUS or luciferase can also be used.

B. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of a desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of a plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly into plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the

construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, et al., *EMBO J.* 3:2717-2722 (1984).

Electroporation techniques are described in Fromm, et al., *Proc. Nat'l. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein, et al., *Nature* 327:70-73 (1987).

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, et al., *Science* 233:496-498 (1984), and Fraley, et al., *Proc. Nat'l. Acad. Sci. USA* 80:4803 (1983).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, et al., *PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE*, pp. 124-176, Macmillian Publishing Company, New York (1983); and Binding, *REGENERATION OF PLANTS, PLANT PROTOPLASTS*, pp. 21-73, CRC Press, Boca Raton (1985). Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, et al., *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

To determine the presence of a reduction or increase of LNP activity, a variety of assays can be used including enzymatic, immunochemical, electrophoretic detection assays (either with staining or western blotting), or complex carbohydrate binding assays.

In a preferred embodiment, a competitive solid phase assay is used to measure LNP activity (Etzler, M.E., *Glycoconj. J.* 11:395 (1994)). This assay measures the ability of various ligands to inhibit the binding of labeled LNP protein to pronase-digested hog gastric mucin blood group A + H substance (HBG A + H) conjugated to Sepharose® (Quinn, J.M. & Etzler, M.E., *Arch. Biochem. Biophys.* 258:535 (1987)).

In another preferred embodiment, an apyrase assay is used to measure LNP activity. See Etzler *et al.*, *Proc. Natl. Acad. Sci. USA* 96:5856-5861 (1999) and Drueckes, P.

et al., *Anal. Biochem.* 230:173-177 (1995). This assay measures the ability of the enzyme to dephosphorylate nucleoside di-and tri-phosphates.

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Asparagus*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucurbita*, *Daucus*, *Glycine*, *Hordeum*, *Lactuca*, *Lycopersicon*, *Malus*, *Manihot*, *Nicotiana*, *Oryza*, *Persea*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Solanum*, *Sorghum*, *Triticum*, *Vitis*, *Vigna*, and *Zea*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Effects of gene manipulation can be observed by northern blots of the mRNA isolated from the tissues of interest. Typically, if the amount of mRNA has increased, it can be assumed that the gene is being expressed at a greater rate than before. Other methods of measuring LNP expression would be by measuring the mycorrhizal infection of the transgenic plants. In addition, levels of LNP could be measured immunochemically, i.e., ELISA, RIA, EIA and other antibody based assays well known to those of skill in the art.

5. Inoculation of Transgenic Plants with Mycorrhizal fungi. Currently, mycorrhizal inoculation of plants can be done on a small scale. Methods used to cultivate mycorrhizal fungi include isolation of spores from soil, pot culture from a single spore to produce colonized root inoculum, and hydroponic or aeroponic methods to grow mycorrhizal plants. Infection of plants with mycorrhizal fungi is know to those of skill in the art. (U.S. Patent 5,096,481, incorporated by reference; Sharma, S. *Microbiologia Sem.* 13:427-436 (1997); Sylvia, D. Vesicular-Arbuscular Mycorrhizal Fungi in *Methods of Soil Analysis Part 2* (1994)). Mycorrhizal cultures can be purchased commercially from International Culture Collection of Arbuscular and Vesicular-Arbuscular Mycorrhizal Fungi (Morgantown, WV, USA) and Tree of Life (San Juan, CA, USA).

6. Examples

The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Characterization of Mycorrhizal infection in the absence of LNP expression.

Generation of stable *Lotus japonicus* transformants that express antisense LNP.

Antisense technology was used in the model legume, *Lotus japonicus*, to test the hypothesis that a novel lectin/nucleotide phosphohydrolase (LNP) is involved in the early events that lead to rhizobial-legume and mycorrhizal-legume symbioses. Antisense lines with no detectable LNP on the surface of the root were unable to undergo root hair deformation and nodulation in response to symbiotic rhizobia. These lines also displayed a severe reduction in ability to colonize with mycorrhizae. The data suggest that in *Lotus japonicus* LNP plays a role in an initial step of both symbioses.

The initiation of the nitrogen-fixing *Rhizobium*-legume symbiosis depends upon specific recognition events that occur between the roots of a particular legume species and the rhizobial strains capable of nodulating that species. Lipochitooligosaccharide signals, called Nod factors (J. Dénarié, F. Debellé, I.-C. Promé, *Ann. Rev. Biochem.* 65:503 (1996); S.R. Long, *Plant Cell* 8:2885 (1996)), produced by the rhizobia have been found to play a key role in this process (P. Mylona, K. Pawlowski, T. Bisseling, *Plant Cell* 7: 869 (1995); J. Cohn, R.B. Day, G. Stacey, *Trends Pl. Sci.* 3:105 (1998)). These signals induce a number of responses in the legume roots that lead to the formation of the root nodules in which the symbioses occur. These responses and the identification of Nod factor binding sites in root membranes (A. Niebel, F. Gressent, J.J. Bono, R. Ranjeva, J. Cullimore, *Biochimie* 81:669 (1999)) imply the existence of Nod factor receptors and an accompanying signal transduction mechanism (J.A. Downie, S.A. Walker, *Current Opinion in Plant Biology* 2:483 (1999)). It has been proposed that some elements of such a plant response mechanism may also play a role in establishing the mycorrhizal-legume symbiosis (V. Gianinaazi-Pearson, *Plant Cell* 8:1871 (1996); M.J. Harrison, *Trends Plant Sci.* 2:54 (1997); C. Albrecht, R. Geurts, T. Bisseling, *EMBO Journal* 18, 281 (1999)). Evidence for this involvement includes the existence of legume mutants that are unable to form either of these symbioses (S.M. Bradbury, R.L. Peterson, S.R. Bowley, *New Phytol.* 124:665 (1993); B. Balaji, A.M. Ba, T.A. LaRue, D. Tepfer, Y. Piché, *Plant Sci.* 102:195 (1994); E. Wegel, L. Schauser, N. Sandal, J. Stougaard, M. Parniske, *MPMI.* 11:933 (1998)) and the induction of early nodulin genes during the colonization of roots by mycorrhizae (C. Albrecht, R. Geurts, T. Bisseling, *EMBO*

Journal **18**, 281 (1999); P. van Rhijn et al., *Proc. Natl. Acad. Sci. U.S.A.* (USA) **94**:5467 (1997)). However, legume genes essential for establishing both of these symbioses have yet to be identified.

A novel lectin with apyrase activity from the roots of the legume, *Dolichos biflorus*, is able to bind Nod factors (M.E. Etzler et al., *Proc. Natl. Acad. Sci. U.S.A.* (USA) **96**:5856 (1999)). This lectin nucleotide phosphohydrolase (LNP) is present on the epidermal cell surface of roots and root hairs in the region of rhizobial attachment. Preincubation of the roots with an antiserum made against recombinant LNP inhibited rhizobial-induced root hair deformation and nodulation. LNP orthologues in other legumes have also been identified. Sequence comparisons with animal and plant apyrases, including other legume apyrases, show that the LNPs appear to constitute a specialized category of apyrases that may be unique to the legumes (N.J. Roberts *et al.*, *Molecular and General Genetics* 262:261(1999)). These correlative studies suggested that LNPs may play a role in the initiation of the *Rhizobium*-legume symbiosis. In the absence of a null mutant to test the possible involvement of LNP in both the *Rhizobium*-legume and *Mycorrhizae*-legume symbioses, antisense technology was utilized in the model legume, *Lotus japonicus* (K. Handberg and J. Stougaard, *Plant J.* **2**: 487(1992); Q. Jiang, P.M. Gresshoff, *MPMI* **10**:59(1997)) to assess the role(s) of this protein.

Three different Lj-LNP antisense constructs were used to generate stable transformants of *L. japonicus*. There was no distinguishable phenotypic growth difference between the wild type and transgenic lines when the plants were grown in the presence of nitrate and ammonia. The ability of the roots to form nodules was tested by inoculating the plants with *Mesorhizobium loti*, a symbiont of *L. japonicus*. Two independent antisense lines, 5'-D-1 and 5'-R-4, formed no nodules with *M. loti*; whereas the other antisense lines and the vector control plants formed healthy nodules (Table 1). Immunoblot analysis of the roots of the transgenic lines revealed that the Nod⁺ transgenic lines generally had wild type levels of a single immunoreactive band of the predicted size of Lj-LNP whereas the two Nod transgenic lines had substantially reduced levels. Confocal immunofluorescence microscopy showed that in contrast to the wild type and the Nod⁺ antisense lines the two Nod lines had no detectable LNP on the surface of the root or root hairs.

Southern blot analyses of the 5'-D-1 and 5'-R-4 Nod⁻ plants showed that both lines have multiple copies of the antisense transgene. Analysis of the 5'-D-1 T1 generation suggests that the expression of the selectable marker (*nptII*) and the Nod phenotype segregate together. Cosegregation of these traits, coupled with the generation of 2 independent Nod

antisense lines, indicates that the Nod phenotype is unlikely to be due to insertional or somatic mutation.

Further examination of the 5'-D-1 Nod⁻ line revealed that after growth in the presence of rhizobia for 4 weeks these plants had numerous root hairs that exhibited a branched or wavy morphology. To investigate the effect of the rhizobia on root hair morphology seedlings were grown in the presence and absence of rhizobia for 4 days (normally a sufficient period of time to see rhizobial-induced root hair deformations). The young and emerging root hairs on the 5'-D-1 Nod plants were almost exclusively straight, whereas typical 'Shepherd's crook' structures were present on both individual and groups of emerging and young root hairs of the wild-type plants. These data suggest that LNP plays a role in the initiation of the *Rhizobium*-legume symbiosis in a process either prior to, and/or including the process of root hair deformation.

Characterization of Mycorrhizal infection of *L. japonicus* lines that express antisense LNP.

To test if LNP is also involved in the mycorrhizal-legume symbiosis, wild type, two of the Nod⁻ lines (5'-D-1 and 5'R) and 3 of the Nod⁺ LNP antisense lines with *Glomus intraradices*, a symbiont of *L. japonicus*. In the 5'-D-1 Nod⁻ line less than 6 % of root segments were infected, as compared to the infection of approximately 89% of root segments in wild type and Nod⁺ transgenic lines (Table 1). On average the shoots of the 5'-D-1 Nod line were 30% shorter than the shoots of the other inoculated plants. The few roots of the 5'-D-1 Nod line that were colonized had a density of vesicles very similar to the wild type. A similar result was reported with *L. japonicus* Myc mutants (E. Wegel, L. Schauser, N. Sandal, J. Stougaard, M. Parniske, *MPMI*. 11:933 (1998)) where the mutants had greatly reduced numbers of colonized roots but the subsequent development of the successful mycorrhizae appeared to be normal. Previously we suggested that LNPs are unique to the legumes (N.J. Roberts *et al.*, *Molecular and General Genetics* 262:261 (1999)); if this hypothesis is correct then either the non-leguminous Myc⁺ plants have a protein that performs a similar function in the mycorrhizal symbiosis or the legumes have uniquely modified the process to involve LNP.

In addition to LNP, legume plants contain other apyrases that more closely resemble the apyrases found in nonleguminous plants. One such conventional apyrase is expressed in *L. japonicus* roots. The relatively high degree of sequence conservation between LNPs and conventional apyrases prompted a comparison of the levels of transcript of this conventional apyrase in the plant lines used in this study. Northern blot analysis

showed the level of this conventional apyrase is approximately equal in wild-type, control and antisense lines, thus establishing that the antisense constructs did not affect the transcript level of this conventional apyrase.

In the absence of detectable LNP on the surface of the root and young and emerging root hairs, *L. japonicus* is Nod⁻ and has greatly reduced mycorrhizal association. This finding suggests that in *L. japonicus* LNP plays a role in both these symbioses. The lack of root hair deformation in the Nod⁻ line after 4 days in the presence of rhizobia indicates that LNP acts upstream of the recently discovered Nin gene, which encodes a reputed transcription factor believed to be involved in the formation of infection threads and initiation of nodule primordia (L. Schauser, A. Roussis, J. Stiller, J. Stougaard, *Nature* 402:191 (1999)).

Table 1

Nodulation and growth data of wild-type and transgenic plants 4 weeks after inoculation with *M. loti*; % of wild-type and transgenic root segments with vesicles 6 weeks after inoculation with *G. intraradices*. Standard deviations are included for all averages.

Line	Generation	Inoculated with <i>M. loti</i>			Inoculated with <i>G. intraradices</i>
		Number of Nodules per Plant ^a	Shoot Mass (mg) per Plant	Root Mass (mg) per Plant	% Root Segments With Vesicles
wild type	F10	23.7 ± 3.2	18.7 ± 4.4	10.2 ± 2.7	87.9 ± 3.0
pBIN19	T1	13.3 ± 2.4	15.0 ± 1.4	8.0 ± 0.4	ND
FL-F-3-1	T2	12.9 ± 1.3	9.5 ± 3.1	5.8 ± 1.4	88.6 ± 1.3
FL-F-3-2	T2	14.8 ± 2.4	13.0 ± 3.8	7.2 ± 1.8	ND
3'-A	T3	12.2 ± 0.7	6.9 ^b ± 1.7	5.1 ^b ± 1.0	ND
3'-B	T3	15.3 ± 3.5	11.3 ± 2.5	6.7 ± 1.4	87.9 ± 3.6
3'-E	T2	20.3 ± 1.7	13.9 ± 4.2	6.8 ± 1.9	ND
5'-D-1	T3	0.0 ± 0.0	8.3 ± 0.2	6.0 ± 0.7	5.5 ± 2.7
5'-D-2	T3	25.5 ± 1.4	20.4 ± 1.8	11.0 ± 0.7	90.6 ± 3.5
5'-H	T2	17.8 ± 1.1	16.9 ± 1.4	8.4 ± 0.4	ND
5'-K	T2	14.5 ± 0.9	10.6 ± 1.9	5.8 ± 1.0	ND
5'-R	T2	0.0 ± 0.0	5.9 ± 1.3	6.1 ± 0.2	5.4 ± 1.4

^aThe average nodule number per plant was calculated using each complete pot as a single replicate. On average the Nod⁺ transgenic lines had fewer nodules than wild-type. It is of interest that a general reduction in nodule number compared to wild-type was also reported in unrelated transgenic kanamycin resistant *Lotus* plants [P. van Rhijn, R.B. Goldberg, A.M. Hirsch, *Plant Cell* 10, 1233 (1998)].

^bThis line was stunted under conditions lacking nitrogen.
ND. Not determined.

METHODS

Construction of LNP antisense vectors. Full-length (FL), 5' and 3' antisense constructs were generated using *Lj-LNP* cDNA (N.J. Roberts et al., *Molecular and General Genetics* 262, 26 I (1999)). These constructs contained nucleotides 1-1489 (FL), nucleotides 1-719 (5') and nucleotides 536-1383 (3'). Each cDNA segment was cloned into the *Xba* I site of the shuttle vector pDH5 1 [J.L. Pietrzak, RD. Shillito, T. Hohn, I. Potrykus, *Nuc. Ac. Res.* **14**, 5857 (1986)] in the reverse orientation relative to the CaMV35S promoter and terminator. The antisense cassette was purified from pDH51 after *Eco*R I digestion and ligated into the *Eco*R I site of the binary vector pBIN19 which utilizes *nptII* as a selectable kanamycin resistance marker under the control of the Nos promoter (M.W. Bevan, *Nuc. Ac. Res.* **12**, 8711(1984)).

Generation of *L. japonicus* lines expressing antisense LNP. The three antisense constructs and a vector control (pBIN19) were transformed into *Agrobacterium tumefaciens*, strain AGL1 (G.L. Lazo, P.A. Stein, R.A. Ludwig, *Bio/Technology* **9**, 963 (1991)), and used to transform *L. japonicus* hypocotyls according to the procedure described by Stiller *et al.*, *J. Exp. Bot.* **48**, 1357 (1997) with the following modifications: a) plants were grown on regeneration medium for only 4-5 days until visible swelling was observed; b) no geneticin was used once the callus was placed on shoot induction medium; and c) full strength cefotaxime was utilized to control agrobacterial growth on all shoot and root media. On some lines successful root initiation was only achieved using the higher auxin concentration in the hairy root regeneration protocol. Regenerated plants were grown under greenhouse conditions and seeds were tested for geneticin resistance on phytigel plates containing 5µg/ml geneticin, 1% sucrose, 1x Gamborg's B5 medium (O.L. Gamborg, RA. Miller, K. Ojima, *Exp. Cell Res.* **50**, 151(1968)). Each of the T0 and subsequent T1 and T2 plants and their seed lines were subjected to Southern blot analyses. Three independent 3' antisense lines (3'-A, B, & E), five independent 5' antisense lines (5'-D1, D2, H. K & R), two FL lines (FL-F-3-1 & 2) and a vector control (pBIN19) were selected for nodulation assays.

Inoculation of *L. japonicus* LNP-antisense lines with *Mesorhizobium loti*.

Four days after germination approximately 50 uniform seedlings from each line were transferred to pots (10x10x8cm) containing sterile vermiculite and perlite (1:1, V/V). The

seedlings were grown for 6 more days and then thinned to approximately 8 plants per pot to maintain uniformity. Four pots from each line were inoculated with 50ml/pot of a fresh overnight culture of *M. loti* diluted to 10^5 cells/ml in Hoagland's solution lacking nitrogen (D.R. Hoagland, D.I. Arnold, Calif Agric. Expt. Sta. Cir. 347, College of Agriculture, Univ. of Calif. Berkeley, CA. (1950)). Preliminary experiments established that this concentration of rhizobia results in approximately a maximum number of nodules/plant. One pot per line was used as an uninoculated control. The distribution of each inoculated pot was randomized in the growth chamber and the plants were grown for 4 weeks using a 26°C, 16 hr day ($350 \mu\text{M m}^{-2} \text{S}^{-1}$), 22°C, 8 hr night. The plants were watered daily using sterile water that was supplemented weekly with sterile Hoagland's solution lacking nitrogen. Four weeks after inoculation the plants were gently removed from the soil, washed and the number of visible nodules were counted on each plant. The shoot and root material in each pot were separated and dried at 65°C for 2 days prior to weighing.

Inoculation of *L. japonicus* LNP antisense lines with *Glomus intraradices*.

The ability to be colonized by the mycorrhizae was tested using 2 independent *Glomus intraradices* sources, International Culture Collection of Arbuscular and Vesicular-Arbuscular Mycorrhizal Fungi (Morgantown, WV, USA) and Tree of Life (San Juan, CA, USA). Inoculum was mixed with sterile vermiculite (1:9 VA'), seeds were germinated on the surface of these mixes in duplicate pots. The plants were watered daily with sterile water that was supplemented weekly with sterile Hoagland's solution 2 containing $20 \mu\text{M NH}_4\text{H}_2\text{PO}_4$, pH 6.8. Plants were grown under similar conditions to the nodulation experiment for approximately 6 weeks. Plants were gently removed from the soil and washed. The roots were separated from the shoots and cleared in 2.5% KOH at 90°C for 90 minutes, rinsed in distilled water then acidified in 1% HCl for 60 minutes prior to staining in Trypan blue at 90°C for 1 hour and destained overnight in 50% glycerol. Roots were cut into approximately 1cm segments and arranged in parallel on glass slides. The presence or absence of vesicles was analyzed using a compound microscope by viewing each segment once over a 2.1 mm field of view.

WHAT IS CLAIMED IS:

1. A method of modulating mycorrhizal infection in a plant, the method comprising introducing into the plant an expression cassette containing a plant promoter operably linked to a heterologous *LNP* polynucleotide or complement thereof, wherein the *LNP* polynucleotide encodes an LNP polypeptide at least about 70% identical to SEQ ID NO:2, SEQ ID NO: 4, or SEQ ID NO: 6
2. The method of claim 1, wherein the heterologous LNP polynucleotide is SEQ ID NO:1.
3. The method of claim 1, wherein the heterologous LNP polynucleotide is SEQ ID NO: 3,
4. The method of claim 1, wherein the heterologous LNP polynucleotide is SEQ ID NO: 5.
5. The method of claim 1, wherein the plant promoter is from an LNP gene.
6. The method of claim 1, wherein the NBP46 polypeptide has an amino acid sequence as shown in SEQ ID NO:2.
7. The method of claim 1, wherein the NBP46 polypeptide has an amino acid sequence as shown in SEQ ID NO: 4.
8. The method of claim 1, wherein the NBP46 polypeptide has an amino acid sequence as shown in SEQ ID NO: 6.
9. The method of claim 1, wherein the expression cassette is introduced into the plant through a sexual cross.
10. The method of claim 1, wherein the promoter is linked to the *LNP* polynucleotide in an antisense orientation.
11. The method of claim 1, wherein the promoter is linked to the *LNP* polynucleotide in a sense orientation.

LNP, A PROTEIN INVOLVED IN THE INITIATION OF MYCORRHIZAL INFECTION IN PLANTS

ABSTRACT OF THE DISCLOSURE

The present invention provides *LNP* polynucleotides that are useful in modulating mycorrhizal infection in plants.

009060" F E 2 2 3 3 5 0

SF 1110124 v1

SEQ ID NO: 1 Complete cDNA sequence of *DB46*

0099060 "T9949960"

```

G A A A C T G A A A C G A G T A C T C T T T C A G T G G T G A G G T T 35
C T G A G A G A T T C A G A A A T G A A T T G G G T G T G G C C A A A 70
G A C A A A G A G C A T G A G C T T C C T A C T C C T C A T C A C T T 105
T T C T A C T C T T C T C A T T G C C A A A A C T T T C T T C T T C G 140
C A A T A T G T T G G G A A C A G T A T C T T A C T A A A T C A T C G 175
T A A G A T A C T T C C C A A C C A G G A A C T C C T T A C C T C T T 210
A C G C T G T C A T C T T T G A T G C T G G T A G C T C T G G G A G T 245
C G T G T C C A T G T C T T C A A T T T T G A C C A G A A C T T A G A 280
T C T C C T G C A C A T T G G C A A T G A C C T C G A G T T T A C A A 315
A A A A G A T C A A A C C C G G T T T G A G C T C A T A C G C T G A T 350
A A G C C T G A A A A A G C T G C A G A A T C T C T C A T T C C A C T 385
T T T G G A G G A A G C T G A A G A T G T T G T C C C T G A G G A A C 420
T G C A C C C C A A G A C A C C C C T T A A G C T T G G G G C A A C A 455
G C A G G T T T G A G G C T C T T G G A T G G G G A T G C T G C T G A 490
A A A G A T A T T G C A A G C G G T T A G G G A A A T G T T C A G G A 525
A C A G A A G T T C C C T G A G C G T T C A A C C T G A T G C A G T A 560
T C T G T T A T T G A T G G A A C C C A A G A A G G T T C T T A C T T 595
A T G G G T T A C A C T T A A A A A C T G T G G G A G T G A T A G A T 630
G A A A G A A G T T T A C A A A A A A C T G T G G G A G T G A T A G A T 665
C T T G G A G G T G C T T C A G T T C A A A T G G C T T A T G C T G T 700
C T C A A G A A A T A C A G C T A A A A A T G C C C C A A A A C C A C 735
C A C A A G G A G A G G A T C C A T A C A T G A A G A A G C T T G T A 770
C T C A A G G G A A A G A A A T A T G A C C T T T A T G T T C A C A G 805
T T A C T T G C G T T A T G G T A A C G A C G C A G C A C G T G T T A 840
A G A T T T T T A A G A C C A C T G A T G G T G C T G C T A G T C C T 875
T G T C T A T T G G C A G G C T A T G A A G A T A T A T A C A G A T A 910
T T C C G G A G A A T C G T A C A A T A T C T A T G G T C C C A C T T 945
C T G G T G C C A A C T T T A A T G A G T G C C G T G A C C T A G C T 980
C T T C A G A T T C T C A G A T T G A A T G A G C C A T G T T C C C A 1015
T G A A A A C T T G C A C C T T T G G T G G G A T A T G G G A T T G G T G 1050
G A A A A G G A A G T G G A C A G A A A A A C C T T G T T G T T A C T 1085
T C A G C T T T C T A C T A T A G G T C T T C T G A G G T T G G T T T 1120
T G T C A C T C C T C C C A A T T C C A A A A A T C G C C C T C T G G 1155
A T T T T T G A A A C T G C A G C T A A A C A A G C T T G T A G T T T A 1190
A C A T T C G A G G A A G C G A A A T C C A C T T T T C C A A A T G T 1225
T G A G A A A G A T A A A C T T C C A T T T G T A T G C G T G G A T T 1260
T C A C A T A C C A G T A T A C A T T G C T T G T T G A T G G A T T T 1295
G G C C T A G A T C C A G A G C A A G A G A T T A C A G T G G C A G A 1330
A G G A A T T G A A T A T C A A G A T G C C A T T G T G G A A A C A G 1365
C A T G G C C T C T A G G A A C T G C C A T A G A A G C C A T A T C A 1400
T C T T T G C C T A A A A T T T A A A T C G T C T A A T G T A T T T A T 1435
C T A A G C C A T G T C C T T C A C T T A T G A C C A C T T T A A T T 1470
A A A A T A A A A C T C A C C C T T T T C A C T A A A A A A A A A A A 1505
A A A A A A A A G T C C T T T T T A T T C C A T T G A G T A T C A A 1540
G T G T T A A T T T G T T T C T G A C A A A T G G A G G T G T A A A A 1575
G T G A A A C A A A G T A T G T T T T T G T C A G A T A C G A A T G G 1610
A A G T A G G G T T A T G A T G A A A A A A A A A A A A A A A A A 1643

```


M N W V P K T K S M S F L L L I T F L L F S L P K L S S Q Y V G N S I L L N -9
H R K I L P N Q E L L T S Y A V I F D A G S S G S R V H V F N F D Q N L D L L H 32
I G N D L E F T K K I K P G L S S Y A D K P E K A A E S L I P L L E F A E D V V 72
P E E L H P K T P L K L G A T A G L R L L D G D A A E K I L Q A V R E M F R N R 112
S S L S V Q P D A V S V I D G T Q E G S Y L W V T V N Y L L G K L G K K F T K T 152
V G V I D L G G A S V Q M A Y A V S R N T A K N A P K P P Q G E D P Y M K K L V 192
L K G K K Y D L Y V H S Y L R Y G N D A A R V K I F K T T D G A A S P C L L A G 232
Y E D I Y R Y S G E S Y N I Y G P T S G A N F N E C R D L A L Q I L R L N E P C 272
S H E N C T F G G I W D G G K G S G Q K N L V V T S A F Y Y R S S E V G F V T P 312
P N S K N R P L D F E T A A K Q A C S L T F E E A K S T P P N V E K D K L P F V 352
C V D F T Y Q Y T L L V D G F G L D P E Q E I T V A E G I E Y Q D A I V E T A W 392
P L G T A I E A I S S L P K F N R L M Y F I 414

```

5'  CAA ATT AAG AAC ATG GAG TTC CTA ATT ACA CTC ATT GCC ACT TTT TTA CTC TTG
    ---
    Q  I  K  N  M  E  F  L  I  T  L  I  A  T  F  L  L  L

      65      74      83      92      101      110
TTA ATG CCT GCA ATC ACT TCC TCC CAA TAT TTA GGA AAC AAC CTA CTC ACT AAT
---
L  M  P  A  I  T  S  S  Q  Y  L  G  N  N  L  L  T  N

      119      128      137      146      155      164
CGA AAG ATT TTC CAA AAA CAA GAA ACC TTA ACC TCT TAC GCT GTC ATA TTT GAT
---
R  K  I  F  Q  K  Q  E  T  L  T  S  Y  A  V  I  F  D

      173      182      191      200      209      218
GCT GGT AGC ACT GGT ACT CGT GTC CAT GTT TAC CAT TTT GAT CAG AAC TTA GAT
---
A  G  S  T  G  T  R  V  H  V  Y  H  F  D  Q  N  L  D

      227      236      245      254      263      272
CTA CTT CAC ATT GGC AAT GAT ATT GAG TTT GTT GAC AAG ATC AAA CCA GGT TTG
---
L  L  H  I  G  N  D  I  E  F  V  D  K  I  K  P  G  L

      281      290      299      308      317      326
AGT GCA TAT GGG GAT AAT CCT GAA CAA GCA GCA AAA TCT CTC ATT CCA CTT TTG
---
S  A  Y  G  D  N  P  E  Q  A  A  K  S  L  I  P  L  L

      335      344      353      362      371      380
GAG GAA GCA GAA GAT GTG GTT CCT GAG GAT CTG CAC CCC AAA ACA CCC CTT AGG
---
E  E  A  E  D  V  V  P  E  D  L  H  P  K  T  P  L  R

      389      398      407      416      425      434
CTT GGG GCA ACC GCA GGT TTG AGG CTT TTG AAT GGG GAT GCT GCT GAA AAG ATA
---
L  G  A  T  A  G  L  R  L  L  N  G  D  A  A  E  K  I

      443      452      461      470      479      488
TTG CAA GCG ACA AGG AAT ATG TTC AGC AAC AGA AGT ACC CTC AAC GTT CAA CGT
---
L  Q  A  T  R  N  M  F  S  N  R  S  T  L  N  V  Q  R

      497      506      515      524      533      542
GAT GCA GTT TCT ATT ATT GAT GGA ACC CAA GAA GGT TCT TAT ATG TGG GTG ACA
---
D  A  V  S  I  I  D  G  T  Q  E  G  S  Y  M  W  V  T

      551      560      569      578      587      596
GTT AAC TAT GTA TTG GGG AAT TTG GGA AAA AGC TTC ACA AAA TCA GTG GGA GTA
---
V  N  Y  V  L  G  N  L  G  K  S  F  T  K  S  V  G  V

      605      614      623      632      641      650
ATT GAC CTT GGA GGT GGT TCA GTT CAA ATG ACA TAT GCA GTG TCA AAG AAA ACA
---
I  D  L  G  G  G  S  V  Q  M  T  Y  A  V  S  K  K  T

```

659	668	677	686	695	704
GCA AAA AAT GCT CCT	AAA GTT GCT	GAT GGA GAG GAT	CCA TAT ATT AAG	AAG CTT	
A K N A P	K V A D	G E D P	Y I K	K L	
713	722	731	740	749	758
GTG CTC AAG GGA AAG	CAA TAT GAT	CTC TAT GTT	CAT AGT TAC	TTG CGT	TTT GGC
V L K G K	Q Y D L	Y V H S	Y L R	F G	
767	776	785	794	803	812
AAA GAA GCA ACT CGA	GCA CAG GTT	TTG AAT GCA	ACT AAT GGA	TCT GCT	AAC CCT
K E A T R	A Q V L	N A T N	G S A	N P	
821	830	839	848	857	866
TGC ATT TTA CCT GGA	TTT AAT GGG	ACC TTT ACA	TAT TCA GGA	GTG GAG	TAT AAG
C I L P G	F N G T	F T Y S	G V E	Y K	
875	884	893	902	911	920
GCT TTT TCC CCT TCT	TCT GGC TCC	AAC TTT GAT	GAT TGC AAA	GAA ATA	ATT CTT
A F S P S	S G S N	F D D C	K E I	I L	
929	938	947	956	965	974
AAG GTT CTT AAA GTA	AAT GAT CCA	TGT CCC TAT	CCG AGT TGC	ACT TTT	GGT GGA
K V L K V	N D P C	P Y P S	C T F	G G	
983	992	1001	1010	1019	1028
ATA TGG AAT GGT GGA	GGA GGG AGT	GGA CAA AAA	AAA CTT TTT	GTT ACT	TCA GCT
I W N G G	G G S G	Q K K L	F V T	S A	
1037	1046	1055	1064	1073	1082
TTC GCT TAC CTG	GCT GAA GAT	GTT GGT ATG	GTT GAG CCA	AAT AAA CCT	AAT TCC
F A Y L A	E D V G	M V E P	N K P	N S	
1091	1100	1109	1118	1127	1136
ATA CTT CAT CCA	GTA GAT TTC	GAA ATT GAA	GCT AAG CGA	GCT TGT GCA	TTA AAC
I L H P V	D F E I E	A K R A	C A L	N	
1145	1154	1163	1172	1181	1190
TTT GAG GAT GTC	AAA TCC ACT	TAT CCT CGA	CTT ACG GAT	GCA AAA CGT	CCA TAT
F E D V K	S T Y P R	L T D A	K R P	Y	
1199	1208	1217	1226	1235	1244
GTA TGC ATG GAT	CTC TTA TAC	CAA CAT GTG	TTG CTT GTT	CAT GGA TTT	GGC TTA
V C M D L	L Y Q H	V L L V	H G F	G L	
1253	1262	1271	1280	1289	1298
GGT CCA CGA AAA	GAG ATT ACA	GTA GGT GAG	GGA ATT CAA	TAT CAG AAT	TCT GTT
G P R K E	I T V G E	G I Q Y	Q N S	V	
1307	1316	1325	1334	1343	1352
GTG GAA GCT GCA	TGG CCT CTA	GGT ACT GCC	GTG GAA GCC	ATA TCA GCG	TTA CCT
V E A A W	P L G T A	V E A I	S A L	P	

1361			1370			1379			1388			1397			1406		
AAG	TTT	AAG	CGA	TTA	ATG	TAT	TTT	ATT	TAA	GCT	TTT	AGA	GAT	GTC	AAG	ATA	TTT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
K	F	K	R	L	M	Y	F	I	*	A	F	R	D	V	K	I	F
1415			1424			1433			1442			1451			1460		
CAG	TAA	CAG	CTA	ACT	TTA	TCA	AAA	ATT	AAA	TAA	AAC	TGG	CGC	ATT	TTG	TCT	TTC
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Q	*	Q	L	T	L	S	K	I	K	*	N	W	R	I	L	S	F

5' AAG TGC TCT TCT CTC TGT AGT TAG TTG CAT TGG ACT AAA GCC ATG GAC TTC TTA

 K C S S L C S * L H W T K A M D F L

ATT AGT CTC ATG ACC TTT GTG TTC ATG TTA ATG CCT GCT ATC TCT TCC TCC CAA

 I S L M T F V F M L M P A I S S S Q

TAT CTC GGA AAC AAC ATT CTC ATG AAT CGT AAG ATA TTA CTC CCC AAA AAT CAG

 Y L G N N I L M N R K I L L P K N Q

GAA CCA GTT ACA TCA TAC GCT GTT ATA TTT GAT GCT GGT ACC ACT GGA AGC AGA

 E P V T S Y A V I F D A G S T G S R

GTC CAT GTC TAC AAT TTT GAT CAG AAC TTA GAT CTC CTT CCC GTT GAA AAC GAA

 V H V Y N F D Q N L D L L P V E N E

CTT GAG TTT TAT GAT TCG GTT AAA CCC GGT TTG AGT TCA TAC GCT GCT AAT CCT

 L E F Y D S V K P G L S S Y A A N P

GAA GAA GCT GCA GAA TCT CTG ATT CCA CTT CTA AAA GAA GCA GAA AAT GTG GTT

 E E A A E S L I P L L K E A E N V V

CCT GTG AGC CAG CAA CCC AAC ACA CCC GTT AAG CTT GGG GCA ACT GCA GGT TTA

 P V S Q Q P N T P V K L G A T A G L

AGG CTT TTG GAG GGG AAT GCT GCT GAA AAT ATA TTG CAA GCG GTC AGG GAT ATG

 R L L E G N A A E N I L Q A V R D M

CTC AGC AAC AGA AGT GCC CTT AAT GTT CAA TCA GAT GCA GTA TCT ATT CTT GAT

 L S N R S A L N V Q S D A V S I L D

GGA ACC CAA GAA GGT TCT TAT CTT TGG GTG ACA ATT AAC TAT CTC TTG GGG AAG

 G T Q E G S Y L W V T I N Y L L G K

TTG GGA AAA AGA TTT ACA AAG ACA GTG GGA GTA GTT GAT CTA GGA GGT GGG TCA

 L G K R F T K T V G V V D L G G G S

009060-12328950

657	666	675	684	693	702
GTG CAA ATG ACA TAT	GCA GTC TCA AGG AAC ACA GCT AAA AAT	GCT CCA AAA GTA			
V Q M T Y	A V S R N T A K N A P K V				
711	720	729	738	747	756
CCT GAA GGA GAG GAT	CCA TAC ATA AAG AAG CTT GTA CTC CAG GGA AAG AAA TAT				
P E G E D	P Y I K K L V L Q G K K Y				
765	774	783	792	801	810
GAC CTT TAT GTT CAC AGT TAC TTG CGC TAT GGA AGA GAA GCA TTT CGT GCA GAG					
D L Y V H S Y L R Y G R E A F R A E					
819	828	837	846	855	864
ATT TTC AAG GTC GCT GGT GGT TCT GCT AAT CCT TGC ATT TTA GCT GGC TTT GAT					
I F K V A G G S A N P C I L A G F D					
873	882	891	900	909	918
GGG GCA TAT ACA TAT TCC GGA GCA GAG TAT AAG GTC TCG GCC CCA GCT TCA GGA					
G A Y T Y S G A E Y K V S A P A S G					
927	936	945	954	963	972
TCT AAC TTG AAT CAA TGC AGA AAG ATA GCT CTT AAG GCT CTT AAA GTG AAT GCA					
S N L N Q C R K I A L K A L K V N A					
981	990	999	1008	1017	1026
CCT TGT CCC TAT CAG AAT TGC ACT TTT GGT GGG ATA TGG AAT GGT GGA GGT GGA					
P C P Y Q N C T F G G I W N G G G G					
1035	1044	1053	1062	1071	1080
AGT GGT CAA AAA AAT CTT TTC CTT ACT TCA TCT TTC TAT TAC CTC TCT GAA GAT					
S G Q K N L F L T S S F Y Y L S E D					
1089	1098	1107	1116	1125	1134
GTT GGG ATC TTT GTG AAT AAA CCC AAT GCC AAA ATT CGT CCA GTT GAT TTG AAG					
V G I F V N K P N A K I R P V D L K					
1143	1152	1161	1170	1179	1188
ACT GCA GCT AAA CTA GCT TGT AAA ACA AAT CTT GAG GAT GCA AAA TCC AAA TAC					
T A A K L A C K T N L E D A K S K Y					
1197	1206	1215	1224	1233	1242
CCA GAT CTT TAT GAG AAA GAC ACT CTT GAA TAT GTG TGC TTG GAT CTT GTC TAC					
P D L Y E K D S V E Y V C L D L V Y					
1251	1260	1269	1278	1287	1296
GTG TAC ACA TTG CTT GTT GAT GGA TTT GGT CTT GAT CCA TTT CAA GAG GTT ACA					
V Y T L L V D G F G L D P F Q E V T					
1305	1314	1323	1332	1341	1350
GTG GCG AAT GAA ATT GAA TAT CAG GAT GCT CTT GTG GAA GCC GCA TGG CCT CTA					
V A N E I E Y Q D A L V E A A W P L					

		1359			1368			1377			1386			1395			1404
GGC	ACT	GCC	ATA	GAA	GCA	ATA	TCA	TCA	TTG	CCT	AAA	TTT	GAG	AGA	TTA	ATG	TAT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
G	T	A	I	E	A	I	S	S	L	P	K	F	E	R	L	M	Y
		1413			1422			1431			1440			1449			1458
TTT	ATT	TAA	ACT	ACT	AGT	ACC	TGC	TTA	AGC	CTG	GAT	TAC	CTG	AAG	AAA	TAA	AAT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
F	I	*	T	T	S	T	C	L	S	L	D	Y	L	K	K	*	N
		1467			1476			1485									
GAA	ATA	AAA	GCC	GCA	TCT	TTC	TTC	CTT	GCT	T	3'						
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
E	I	K	A	A	S	F	F	L	A								